



ECOLE NATIONALE SUPÉRIEUR PARIS-SACLAY
MASTER MATHÉMATIQUES VISION APPRENTISSAGE

TP 1: Analyse Spectrale et Inférence

OREISTEIN Pierre

COURS: THÉORIE DES MATRICES ALÉATOIRES

APRIL 10, 2019

1 - Séparation exacte du spectre:

1.1

On considère la loi normale comme loi ayant un moment d'ordre 4 fini pour les entrées de la matrice X . On génère alors plusieurs matrices de covariances, on calcule leur spectre et on représente la distribution des valeurs propres des spectres obtenus.

En choisissant une matrice R_N ayant pour valeurs propres 1, 3 ou 7 avec probabilités 0.2, 0.3 ou 0.5 respectivement, et pour dimensions $N = 20$ et $n = 1000$ (*i.e.* $c = 0.02$), on obtient le graphique ci-dessous:

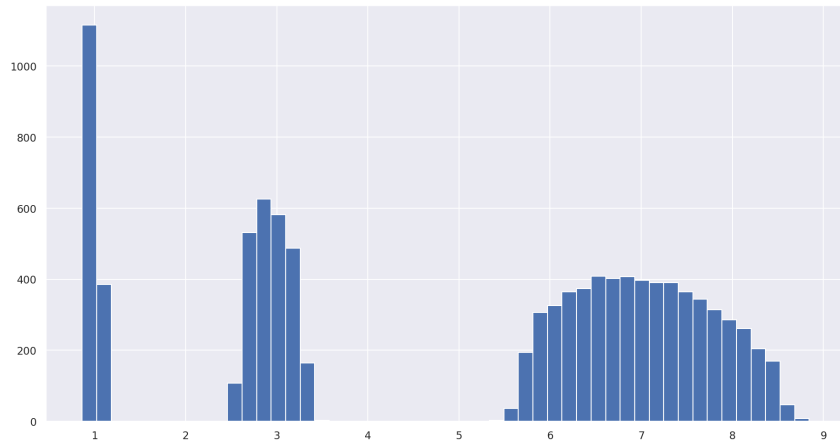


Figure 1 – Distribution des valeurs propres du spectre de 500 matrices de covariances générées à partir de R_N avec une loi normale pour les entrées de X .

On retrouve alors les trois résultats attendus:

- Pas de valeur propre qui s'éloigne aux infinis.
- Séparation exacte du support de chaque valeur propre.
- Le nombre de valeurs propres dans chaque composante connexe est proportionnel à N_k .

1.2

On considère cette fois-ci le cas où la loi n'est plus de moment d'ordre 4 fini. On choisit cette fois-ci la loi de Student. En reprenant les mêmes paramètres qu'à la question précédente, on obtient la figure ci-dessous:

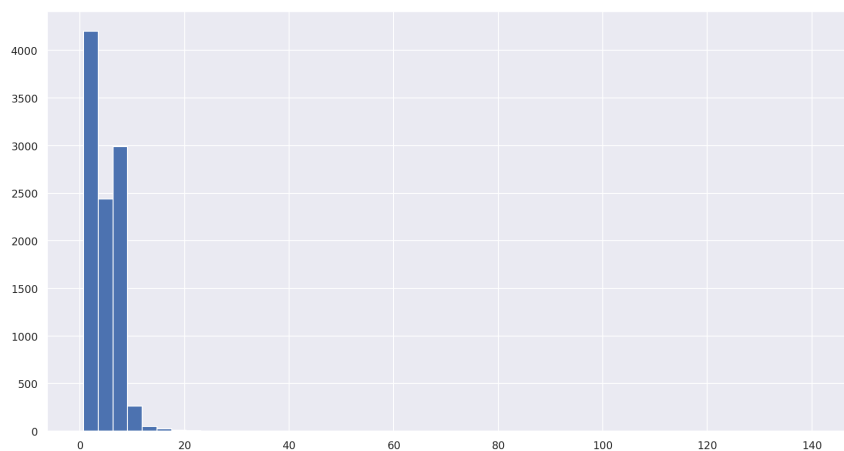


Figure 2 – Distribution des valeurs propres du spectre de 500 matrices de covariances générées à partir de R_N avec une loi de Student pour les entrées de X . Valeur maximale (non visible): 141

Cette fois, on observe que:

- Certaines valeurs propres tendent vers l'infini, ou du moins, sont beaucoup plus grande qu'attendues. Sur la figure ci-dessus, le maximum n'est pas visible mais vaut 141.
- Il n'y a plus de séparation exacte du spectre.

2 Graphe de $x(\tilde{t})$

2.1

Comme vu en cours, $\tilde{t}(z)$ est la transformée de Stieljes (TS) de la limite des mesures empiriques des valeurs propres des matrice de covariance $\frac{1}{n}X_N^*R_NX_N$. De plus, $\tilde{t}(z)$ vérifie l'équation suivante:

$$x(\tilde{t}) = \frac{-1}{\tilde{t}} + c \int_{\mathbb{R}} \frac{u}{1 + u\tilde{t}} L_{\infty}^R(du), \quad \forall \tilde{t} \in \mathbb{R} \setminus \left\{ \tilde{t} \mid \frac{-1}{\tilde{t}} \in \text{Supp}(L_{\infty}^R) \right\} \quad (1)$$

De plus, ici on connaît R_N . En inversant l'équation précédente, on obtient donc:

$$\tilde{t} = \frac{1}{-(x + \mathbf{i}y) + c \sum_{k=1}^K \frac{N_k}{N} \frac{\lambda_k}{1 + \lambda_k \tilde{t}}}$$

On peut alors utiliser l'algorithme du point fixe pour approximer \tilde{t} pour tout x et ensuite tracer le graphe de $f(x) = \frac{1}{\pi} \lim_{y \rightarrow 0} \Im(\tilde{t}(x + \mathbf{i}y))$. On obtient alors le graphe ci-dessous:

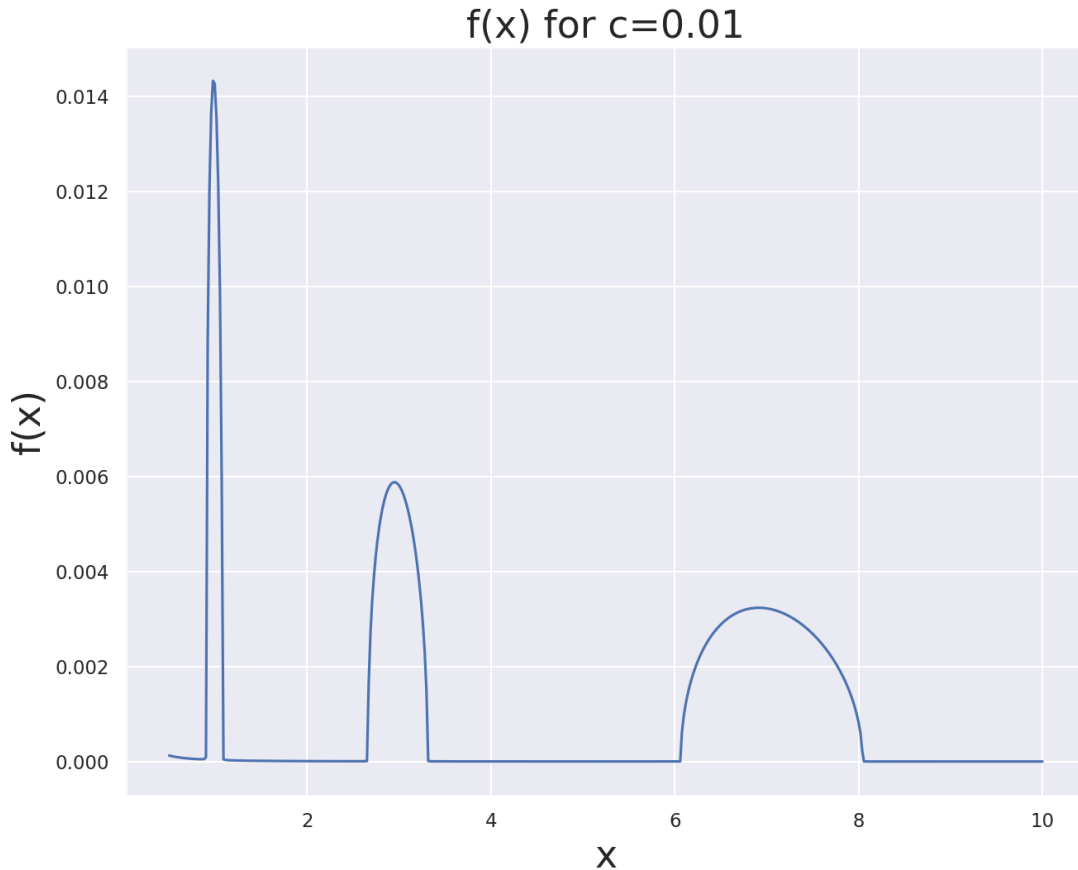


Figure 3 – Graphe de la fonction $f(x)$

On retrouve la loi de l'histogramme observée en question 1.1.

2.2

De même, en réutilisant l'équation 1, on peut tracer $x(\tilde{t})$, $\forall \tilde{t} \notin \left\{ \frac{-1}{\lambda_1^R}, \dots, \frac{-1}{\lambda_K^R} \right\}$:

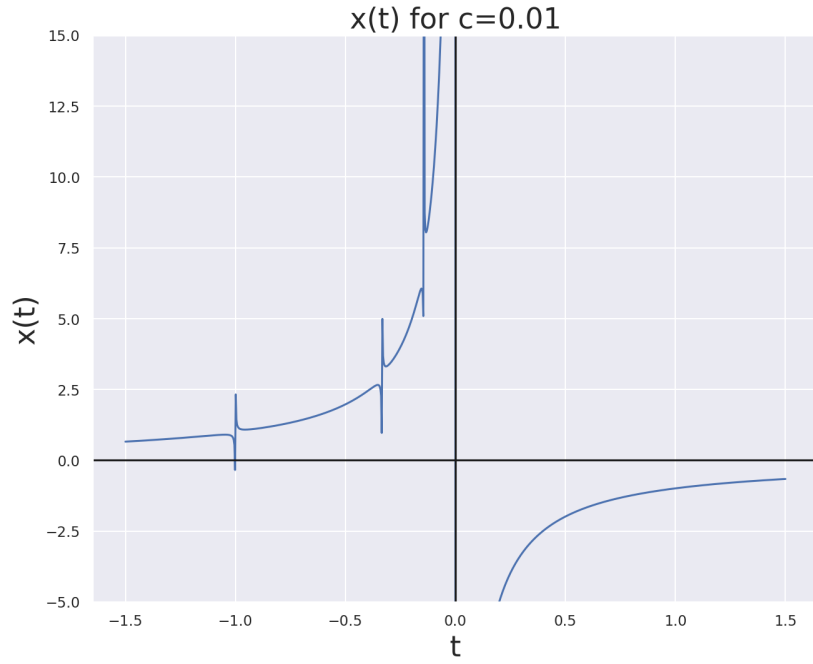


Figure 4 – Graphe de la fonction $x(\tilde{t})$

On retrouve bien les asymptotes verticales aux voisinages des $-1/\lambda_k^R$.

2.3

Dans cette section, on va étudier l'influence du choix de c , $(c_i)_{i=1\dots K}$ et $(\lambda_k)_{k=1\dots K}$ sur le graphe de f et de x .

Influence de c

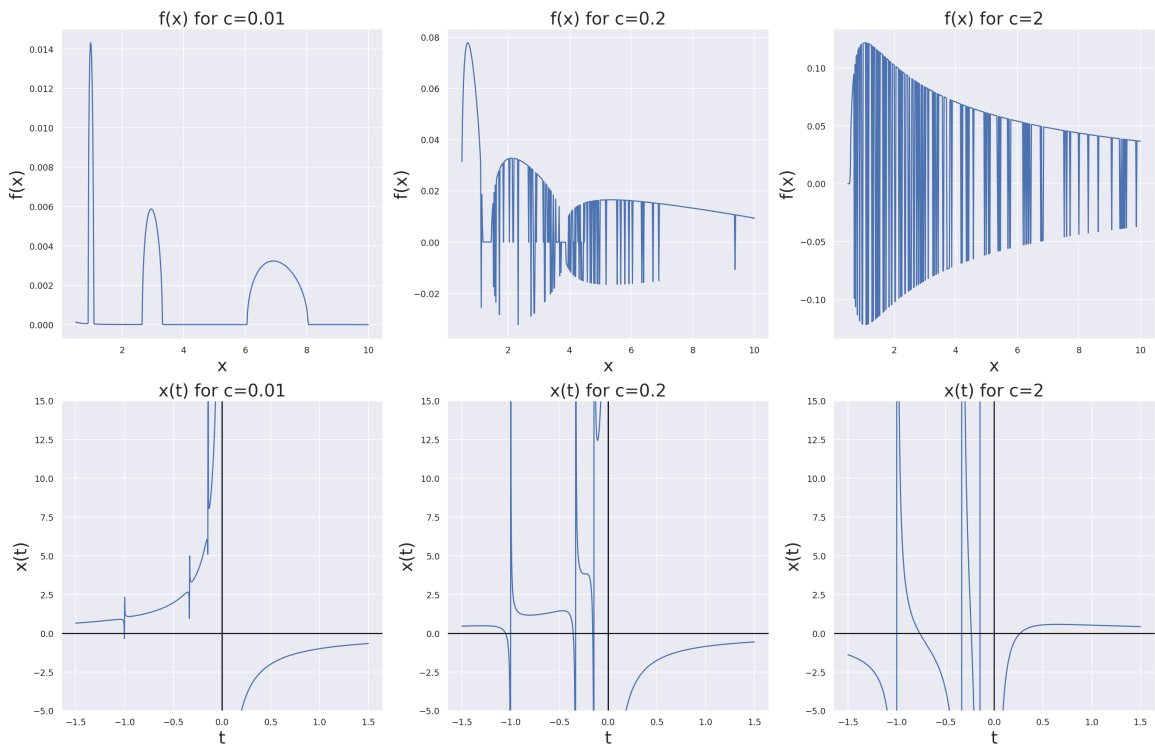


Figure 5 – Influence du choix de c sur les graphes de f et de x .

Ci-dessus sont représentés sur la première ligne le graphe de f et sur la seconde ligne le graphe de x pour des valeurs de c égales à 0.01, 0.05 et 2 respectivement.

- Quand c est faible (cas de $c=0.01$, première colonne), les graphes de f et de x correspondent à ce qui étaient attendus: séparation du spectre en particulier pour le graphe de f .
- Quand c augmente (cas de $c=0.2$), on constate que du "bruit" apparaît par rapport aux graphes précédent. De plus, la distinction entre les clusters est moins nette. Enfin, les parties croissantes du graphe de x sont plus étroites. Il y a donc une réduction du support de la mesure limite associée.
- Finalement, quand c est grand ($c=2$), on constate que le graphe de f ne sépare plus les clusters. De plus, le graphe de x admet toujours les mêmes asymptotes mais la croissance de cette fonction est différente ce qui traduit une modification du support de la mesure limite associée.

Influence de $(c_i)_{i=1\dots K}$

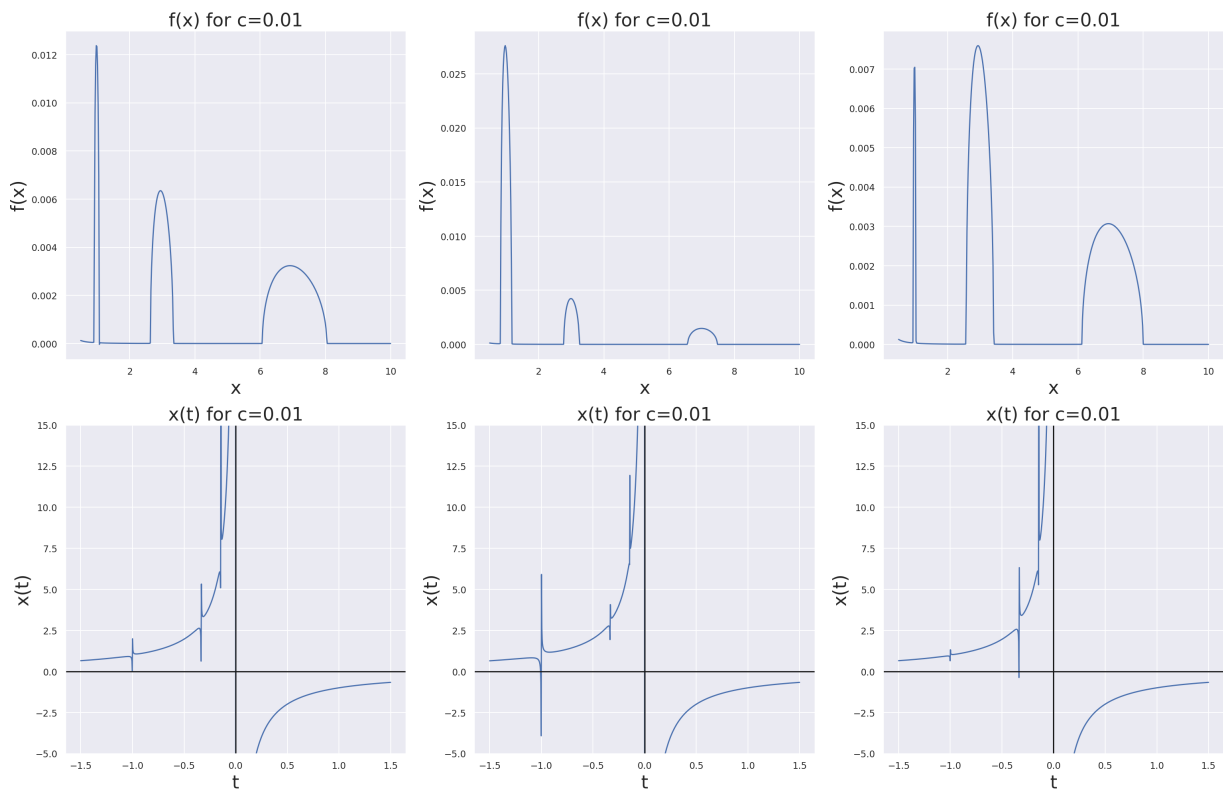


Figure 6 – Influence du choix de $(c_i)_{i=1\dots K}$ sur les graphes de f et de x .

Ci-dessus sont représentés les graphes de f (première ligne) et de x (deuxième ligne). On étudie ici l'influence du choix des c_i sur ces graphes. On a pris les mêmes λ_k que précédemment (1, 3 et 7) mais on représente les graphes de f et x pour différents choix de c_i : $[0.15, 0.35, 0.5]$, $[0.73, 0.16, 0.11]$ et $[0.05, 0.5, 0.45]$ respectivement)

- *Influence sur le graphe de f* : Comme vu précédemment, on retrouve bien le fait que les parties connexes liés à une valeur propre sont proportionnel à c_i . De même, lorsque qu'un λ_i est associé à un c_i faible, l'intervalle de sa partie connexe associé est peu étendu (e.g: partie connexe associé à 7 entre les colonnes 1 et 2).
- *Influence sur le graphe de x* : Lorsqu'on a un c_i élevé, on constate une meilleur approximation au niveau de l'asymptote associée à λ_i (asymptote associé à $\lambda_i = 1$ entre la première ($c_1 = 0.15$) et deuxième colonne ($c_1 = 0.73$)).

Influence de $(\lambda_k)_{k=1\dots K}$

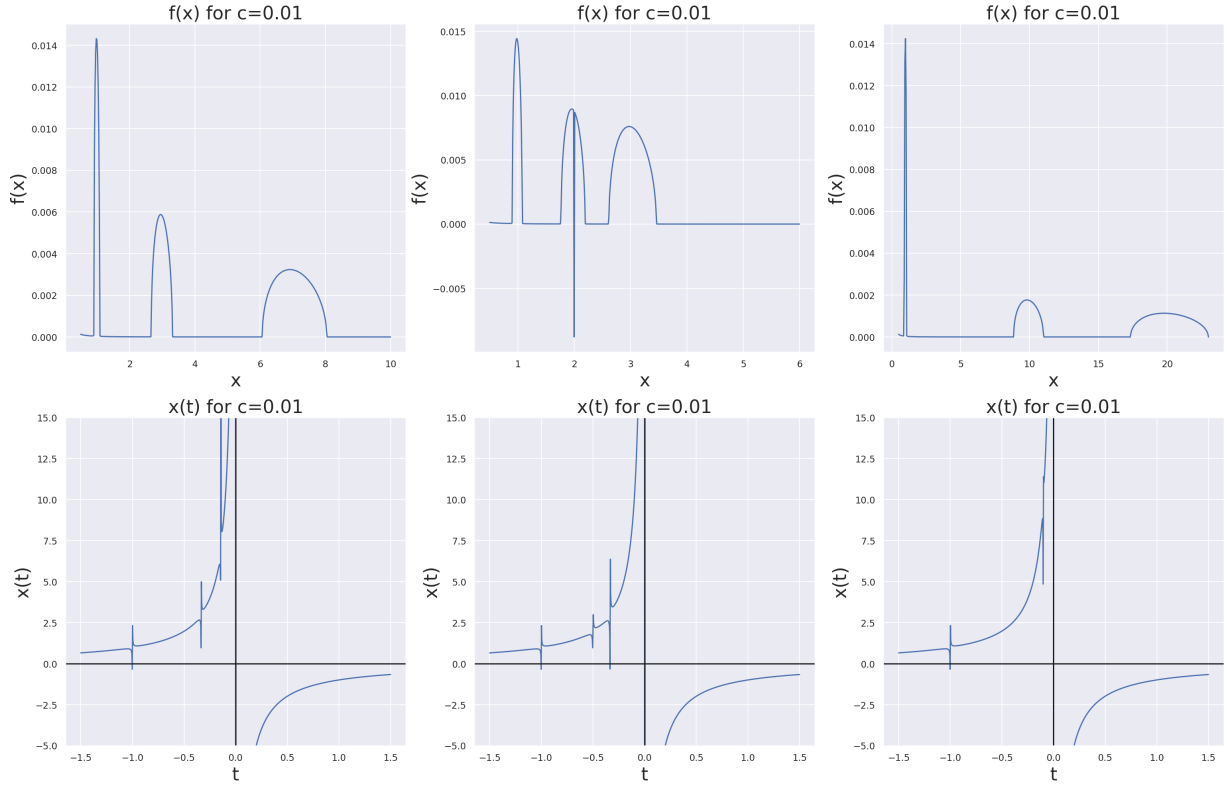


Figure 7 – Influence du choix de $(\lambda_k)_{k=1\dots K}$ sur les graphes de f et de x .

Ci-dessus sont représentés les graphes de f (première ligne) et de x (deuxième ligne). On étudie ici l'influence du choix des λ_k . On a pris les mêmes c_i que précédemment (0.2, 0.3 et 0.5 respectivement) mais on représente les graphes de f et x pour différents choix de λ_k : $([1, 3, 7], [1, 2, 3]$ et $[1, 10, 20]$ respectivement).

- *Influence sur le graphe de f* : On constate que plus un λ_k est grand, plus sa partie connexe associé est étendue (e.g: partie connexe associé à $\lambda_i = 20$ dans la dernière colonne).

3 Estimation des λ_i^R

3.1

Grâce aux simulations réalisées pour la partie séparation exacte de spectre 1, on sait que l'ensemble de paramètres ($N=20$, $n=1000$, $nb=500$, $law="normal"$, $lambdas=[1, 3, 7]$) est un ensemble qui satisfait la séparation exacte du spectre.

3.2

(a) En reprenant les calculs fait en cours et en choisissant $f(u) = u$ et un contour quelconque Γ , on a:

$$\begin{aligned}
 \frac{1}{N} \sum_{i=1 \text{ s.t. } \lambda_i^R \in \Gamma} \lambda_i^R &\simeq \frac{-1}{2\pi i} \oint_{\Gamma} \frac{n}{N} (-w \tilde{g}_n(w) + \frac{N}{n} - 1) \frac{\tilde{g}_n'(w)}{\tilde{g}_n(w)^2} dw \\
 &\simeq \frac{-1}{2\pi i} \oint_{\Gamma} \frac{wn}{N} \frac{\tilde{g}_n'(w)}{\tilde{g}_n(w)} dw - \frac{1}{2\pi i} \oint_{\Gamma} (\frac{n}{N} - 1) \frac{\tilde{g}_n'(w)}{\tilde{g}_n(w)^2} dw \\
 &\simeq \frac{-1}{2\pi i} \oint_{\Gamma} \frac{wn}{N} \frac{\tilde{g}_n'(w)}{\tilde{g}_n(w)} dw - \frac{1}{2\pi i} (\frac{n}{N} - 1) \underbrace{\left[\frac{1}{\tilde{g}_n(w)} \right]_{\Gamma}}_{=0}
 \end{aligned}$$

Or de plus, on a: $\tilde{g}_n(w) = \frac{1}{n} \sum_{i=1}^n \frac{1}{w-\lambda_i}$ On obtient donc:

$$\sum_{i=1, s.t. \lambda_i^R \in \Gamma}^N \lambda_i^R = \frac{n}{2\pi i} \oint_{\Gamma} w \underbrace{\frac{\sum_{i=1}^n \frac{1}{(w-\lambda_i)^2}}{\sum_{i=1}^n \frac{1}{w-\lambda_i}}}_{=h(w)} dw \quad (2)$$

Pour résoudre cette intégrale, on va utiliser le théorème des résidus et on va donc chercher les pôles de h . Vu que \tilde{g}_n est au dénominateur, on a donc que les zéros de \tilde{g}_n sont des pôles de h . Notons-les $(\eta_i)_{i=1\dots n}$ comme en cours. De plus, on peut réécrire h :

$$\begin{aligned} h(w) &= w \frac{\sum_{i=1}^n \frac{1}{(w-\lambda_i)^2}}{\sum_{i=1}^n \frac{1}{w-\lambda_i}} \\ &= w \frac{\sum_{i=1}^n \prod_{j \neq i} (w-\lambda_j)^2}{\sum_{i=1}^n (w-\lambda_i) \prod_{j \neq i} (w-\lambda_j)^2} \\ &= w \frac{\sum_{i=1}^n \prod_{j \neq i} (w-\lambda_j)^2}{\prod_{j=1}^n (w-\lambda_j) \sum_{i=1}^n \prod_{j \neq i} (w-\lambda_j)} \end{aligned}$$

On en déduit que les $(\lambda_i)_{i=1\dots n}$ sont également des pôles de h . On a donc trouver tous les pôles. On peut donc utiliser le théorème des résidus. Pour rappel:

Theorem 3.1 (Théorème des Résidus)

Si $(a_k)_{k=1\dots K}$ sont les pôles de h dans Γ alors:

$$\frac{1}{2\pi i} \oint_{\Gamma} h(w) dw = \sum_{k=1}^K \lim_{w \rightarrow a_k} (w - a_k) h(w)$$

En utilisant ce théorème, on obtient donc:

$$\sum_{i=1, s.t. \lambda_i^R \in \Gamma}^N \lambda_i^R = n \left(\sum_{i=1, s.t. \eta_i \in \Gamma}^n \lim_{w \rightarrow \eta_i} (w - \eta_i) h(w) + \sum_{i=1, s.t. \lambda_i \in \Gamma}^n \lim_{w \rightarrow \lambda_i} (w - \lambda_i) h(w) \right)$$

Il ne nous reste plus qu'à utiliser la règle de l'hôpital:

Lemma 3.1 (Règle de l'Hôpital)

Si f et g sont deux fonctions définies au voisinage de a , dérivable en a et telles que $f(a) = g(a) = 0$ et $g'(a) \neq 0$, alors:

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}$$

On a donc:

- Pour $a = \eta_i$:

$$\begin{aligned} \lim_{w \rightarrow \eta_i} (w - \eta_i) h(w) &= \lim_{w \rightarrow \eta_i} \frac{(w - \eta_i) w \sum_{i=1}^n \frac{1}{(w-\lambda_i)^2}}{\sum_{i=1}^n \frac{1}{(w-\lambda_i)}} \\ &= -\eta_i \end{aligned}$$

- Pour $a = \lambda_i$, on n'a même pas besoin de la règle de l'Hôpital:

$$\begin{aligned} \lim_{w \rightarrow \lambda_i} (w - \lambda_i) h(w) &= \lim_{w \rightarrow \lambda_i} \frac{(w - \lambda_i) w \sum_{j=1}^n \frac{1}{(w-\lambda_j)^2}}{\sum_{k=1}^n \frac{1}{(w-\lambda_k)}} \\ &= \lim_{w \rightarrow \lambda_i} \frac{w \left(1 + \sum_{j \neq i}^n \frac{(w-\lambda_i)^2}{(w-\lambda_j)^2} \right)}{1 + \sum_{j \neq i}^n \frac{(w-\lambda_i)}{(w-\lambda_j)}} \\ &= \lambda_i \end{aligned}$$

Ainsi au vu de ces deux résultats et en utilisant un contour Γ_k qui entoure λ_k on en déduit:

$$N_k \lambda_k^R = n \left(\sum_{i=1, s.t. \eta_i \in \Gamma_k}^n -\eta_i + \sum_{i=1, s.t. \lambda_i \in \Gamma}^n \lambda_i \right)$$

Ainsi, si on utilise cette équation pour approximer les $(\lambda_k)_{k=1\dots K}$, on trouve une erreur inférieure à moins de 1% pour l'estimation de chacun des λ pour un ensemble de paramètres égales à: $N = 20$, $n = 1000$, $\lambda_k \in \{1, 3, 7\}$ avec probabilités respectives 0.2, 0.3 et 0.5 respectivement.

(b) On peut alors étudier la loi de ces estimateurs $\hat{\lambda}_k^R$. En particulier, on peut tracer l'histogramme de ces estimateurs pour les mêmes paramètres que ceux de la question précédente. On obtient alors le graphe ci-dessous:

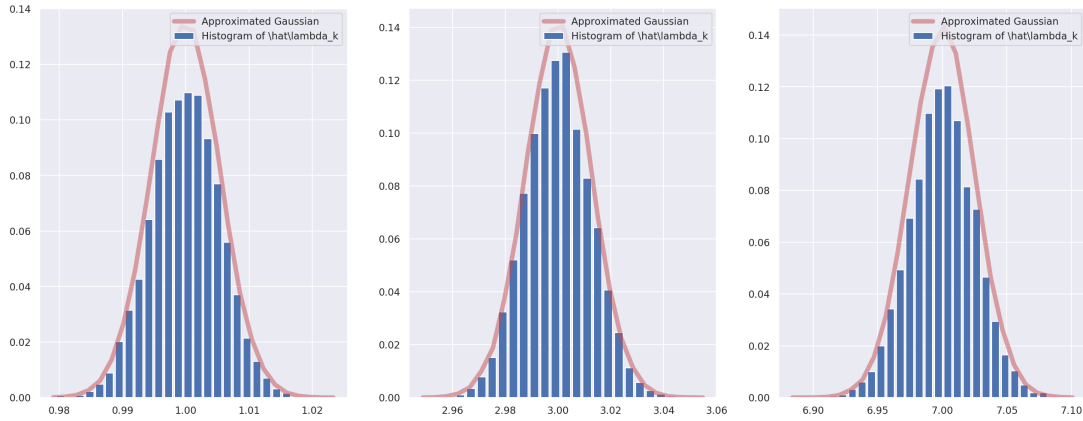


Figure 8 – Distribution empirique des estimateurs pour chacun des $\lambda_k \in \{1, 3, 7\}$ (en bleu) accompagné d'une distribution gaussienne de même moyenne et variance empirique que les données. (en orange)

Des figures ci-dessus, on peut proposer l'hypothèse que les estimateurs des λ_k suivent une loi gaussienne de moyenne λ_k . La variance de ces lois est difficile à déterminer directement, mais elle semble dépendre de manière croissante des λ_k .

(c) Au vu de ces premiers résultats, on peut évaluer l'erreur quadratique (MSE) en fonction de N des $\hat{\lambda}_k^R$. On obtient alors la figure suivante:

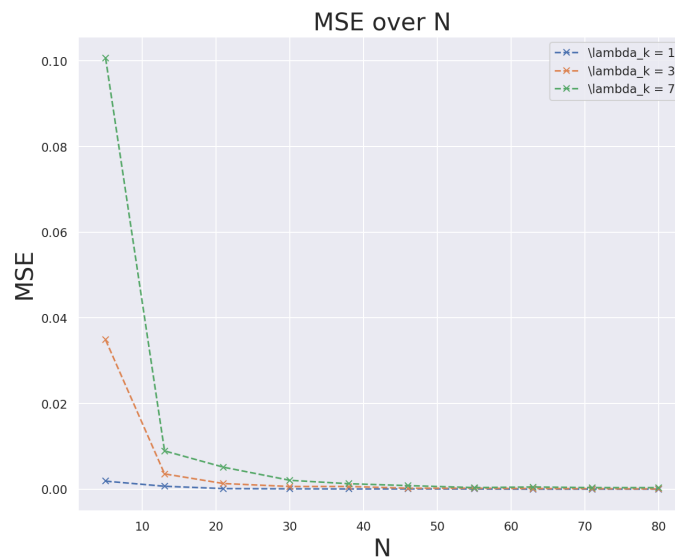


Figure 9 – MSE en fonction de N pour chaque estimateur $\hat{\lambda}_k^R$ pour $\lambda \in \{1, 3, 7\}$.

Grâce à cette nouvelle figure, on en déduit que l'erreur MSE des estimateurs $\hat{\lambda}_k^R$ décroît avec N et aussi, comme observé précédemment, dépendent de manière croissante de λ_k .

(d) Dans un second temps, on peut s'intéresser au cas d'une estimation naïve des λ_k pour mesurer les potentiels avantages de la méthode précédente. On trace donc de nouveau la MSE en fonction de N avec des estimateurs naïfs (moyenne des λ_i dans chaque cluster). On obtient alors le graphique suivant pour le même ensemble de paramètres que celui des question précédentes:

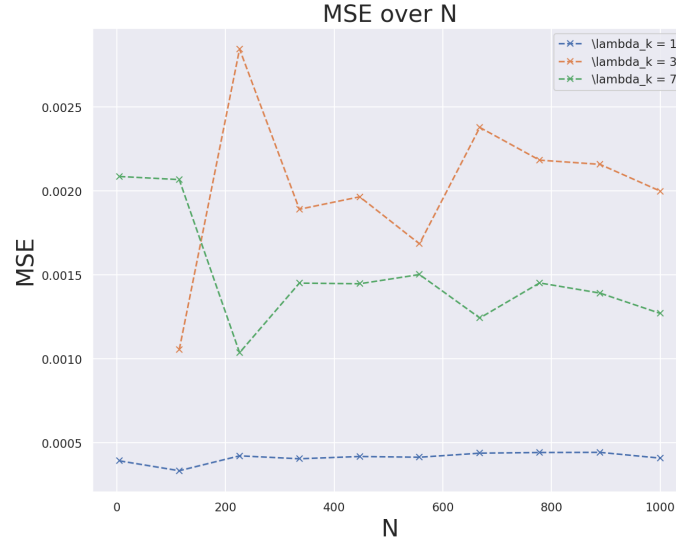


Figure 10 – MSE en fonction de N pour des estimateurs naïfs de $\lambda \in \{1, 3, 7\}$.

On constate que l'erreur ne décroît plus avec N pour ces estimateurs naïfs. On en déduit donc que ceux-ci ne sont pas consistants. De même, on constate que cette erreur croît avec la valeur de λ_k . Néanmoins, l'erreur reste faible et ainsi l'estimateur naïf peut être utile en pratique du fait de sa simplicité de calcul.

3.3

On se place désormais dans le cas où $K = 2$. Après diverses simulations, on constate qu'une paramétrisation proche de $c = 0.2 = c_0$ correspond à un régime critique. En effet, pour des valeurs légèrement inférieure, on observe une séparation en deux clusters tandis que pour des valeurs légèrement supérieure, on n'observe plus qu'un seul cluster. Ci-dessous est représenté l'histogramme des valeurs propres pour différentes matrices de covariances générées aléatoirement à partir de R pour $\lambda_k \in \{1, 3\}$ apparaissant avec probabilités 0.4 et 0.6 respectivement:

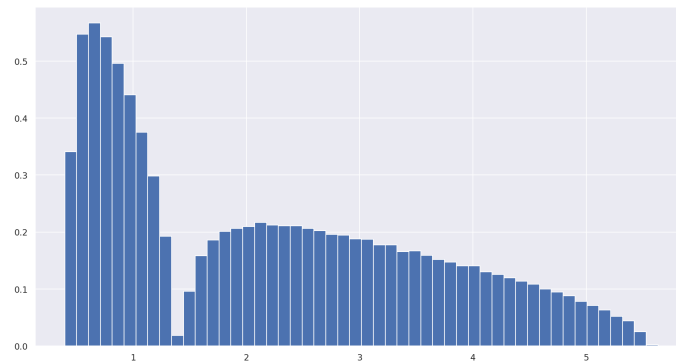


Figure 11 – Distribution des valeurs propres de matrices de covariance générées aléatoirement pour $\lambda_k \in \{1, 3\}$ dans le cas d'un régime critique: $c = c_0 = 0.2$.

3.4

(a) On s'intéresse maintenant au comportement des estimateurs $\hat{\lambda}_k^R$ au voisinage du point critique $c = c_0$. On trace donc l'erreur MSE en fonction de n pour $N = 20$. On choisit des valeurs de n telle que c_0 soit contenu dans l'intervalle considéré pour n . En prenant en compte ces pré-requis, on obtient le graphe ci-dessous:



Figure 12 – Erreur d'estimation MSE des estimateurs $\hat{\lambda}_i^R$ au voisinage du point critique $c = c_0 = 0.2$ pour $N = 20$ en fonction de n .

Grâce à cette figure, on constate tout d'abord que l'erreur MSE a fortement augmenté en comparaison avec les régimes précédents où c était plus faible 9. De même, l'estimation devient très instable. En fait, comme on utilise un contour pour isoler chaque cluster, lorsque ceux-ci fusionnent ($c \leq c_0$), l'approximation par cluster individuel devient peu fiable.

(b) Du fait du manque de séparabilité, on ne peut plus évaluer les valeurs λ_k de manière individuelles. Néanmoins, en utilisant $f(u) = u$ et $f(u) = u^2$, on va pouvoir estimer $(N_1\lambda_1 + N_2\lambda_2)/N$ et $(N_1\lambda_1^2 + N_2\lambda_2^2)/N$ puis résoudre ce système pour obtenir les valeurs individuelles. En pratique, on va donc utiliser le résultat vu plus haut ainsi que celui du cours:

$$N_1\lambda_1^R + N_2\lambda_2^R = n \left(\sum_{i=1, s.t. \eta_i \in \Gamma}^n -\eta_i + \sum_{i=1, s.t. \lambda_i \in \Gamma}^n \lambda_i \right) = e_1$$

$$N_1(\lambda_1^R)^2 + N_2(\lambda_2^R)^2 = n \sum_{i=1, s.t. \eta_i \in \Gamma}^n \frac{1}{\frac{1}{n} \sum_{j=1}^n \frac{1}{(\lambda_j - \eta_i)^2}} = e_2$$

En choisissant Γ un contour contenant λ_1 et λ_2 .

On peut maintenant résoudre le système d'équations:

$$\begin{aligned}
\begin{cases} N_1\lambda_1^R + N_2\lambda_2^R &= e_1 \\ N_1(\lambda_1^R)^2 + N_2(\lambda_2^R)^2 &= e_2 \end{cases} &= \begin{cases} \lambda_2^R &= \frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \\ (\lambda_1^R)^2 + \frac{N_2}{N_1} \left(\frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \right)^2 &= e_2 \end{cases} \\
&= \begin{cases} \lambda_2^R &= \frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \\ (\lambda_1^R)^2 + \frac{N_2}{N_1} \left(\left(\frac{e_1}{N_2} \right)^2 - 2 \frac{N_1 e_1 \lambda_1^R}{N_2^2} + \left(\frac{N_1 \lambda_1^R}{N_2} \right)^2 \right) &= \frac{e_2}{N_1} \end{cases} \\
&= \begin{cases} \lambda_2^R &= \frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \\ (\lambda_1^R)^2 \left(1 + \frac{N_1}{N_2} \right) - 2 \frac{e_1}{N_2} \lambda_1^R + \frac{(e_1)^2}{N_1 N_2} &= \frac{e_2}{N_1} \end{cases} \\
&= \begin{cases} \lambda_2^R &= \frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \\ (\lambda_1^R)^2 - 2 \frac{e_1}{(N_1+N_2)} \lambda_1^R + \frac{(e_1)^2}{N_1(N_1+N_2)} &= \frac{e_2 N_2}{N_1(N_1+N_2)} \end{cases} \\
&= \begin{cases} \lambda_2^R &= \frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \\ \left(\lambda_1^R - \frac{e_1}{(N_1+N_2)} \right)^2 - \left(\frac{e_1}{(N_1+N_2)} \right)^2 + \frac{(e_1)^2}{N_1(N_1+N_2)} &= \frac{e_2 N_2}{N_1(N_1+N_2)} \end{cases} \\
&= \begin{cases} \lambda_2^R &= \frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \\ \left(\lambda_1^R - \frac{e_1}{(N_1+N_2)} \right)^2 &= \frac{e_2 N_2}{N_1(N_1+N_2)} + (e_1)^2 \left(\frac{N_1 - (N_1+N_2)}{N_1(N_1+N_2)^2} \right) \end{cases} \\
&= \begin{cases} \lambda_2^R &= \frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \\ \left(\lambda_1^R - \frac{e_1}{(N_1+N_2)} \right)^2 &= \frac{1}{N_1(N_1+N_2)^2} ((-N_2)(e_1)^2 + N_2(N_1 + N_2)e_2) \end{cases} \\
&= \begin{cases} \lambda_2^R &= \frac{e_1}{N_2} - \frac{N_1\lambda_1^R}{N_2} \\ \lambda_1^R &= \frac{e_1}{(N_1+N_2)} \pm \sqrt{\frac{1}{N_1(N_1+N_2)^2} ((-N_2)(e_1)^2 + N_2(N_1 + N_2)e_2)} \end{cases}
\end{aligned}$$

Dès lors se pose un problème du choix du signe devant la racine.

En implémentant la procédure décrite ci-dessus et en choisissant le signe adapté pour un résultat cohérent de λ_1^R , on trouve une erreur inférieure à 3% dans l'estimation des deux λ_k^R dans le cadre d'un ensemble de paramètres égale à ($N = 20$, $n = 100$, (*i.e.* $c = 0.2$) et $\lambda \in \{1, 3\}$ avec probabilité d'apparition égale à 0.4 et 0.6 respectivement).

(c) En utilisant la méthode d'estimation proposé dans la question précédente, on va re-tracer l'erreur MSE d'estimation des λ_k^R pour ces nouveaux estimateurs dans le voisinage de c_0 . On obtient alors la figure suivante:

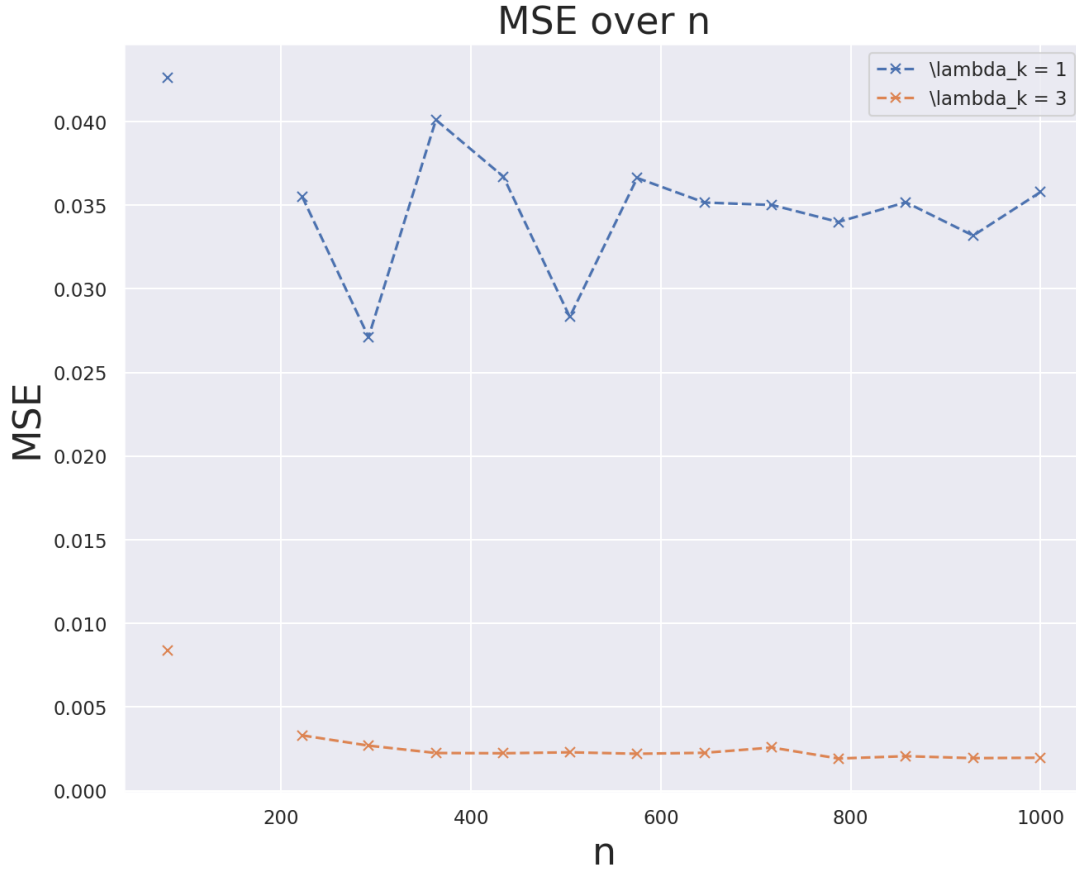


Figure 13 – Erreur d'estimation MSE des estimateurs λ_i^R au voisinage du point critique $c = c_0 = 0.2$ pour $N = 20$ en fonction de n .

On constate cette fois-ci une moins grande instabilité au voisinage du point critique mais surtout une erreur MSE beaucoup plus faible que sur la figure précédente pour les estimateurs individuels (cf FIGURE 12). L'absence de valeurs pour $n = 100$ s'explique par une erreur d'approximation numérique que nous ne sommes pas parvenu à résoudre.

(d) Cette méthode bien qu'efficace présente cependant deux grandes limitations:

- La non unicité de la solution du système d'équation précédent. Le processus de détection n'est donc pas totalement automatique.
- Une efficacité seulement dans le cas $K = 2$. Dès lors, il semble peu envisageable de continuer une telle procédure pour des valeurs de K supérieures. En effet, il faudrait faire de nouveau calculs pour un nouveau f pour faire apparaître une nouvelle équation linéairement indépendante des précédentes.

Une possible piste de recherche serait sur le choix de la fonction f lorsque l'on calcul l'intégrale comme en question 3.2. En effet, on aimerait trouver f_θ possiblement paramétré par un θ telle que le calcul de l'intégrale soit possible pour tout θ et générant un système d'équations linéairement indépendants admettant une unique solution pour le choix de plusieurs θ et ce pour un grand nombre de valeurs de K .